(12) UK Patent Application (19) GB (11) 2 343 822 (13) A

(43) Date of A Publication 17.05.2000

(21) Application No 0001586.7

(22) Date of Filing 01.07.1998

Date Lodged 24.01.2000

(30) Priority Data
(31) 9714001 (32) 02.07.1997 (33) GB

(62) Divided from Application No 9814279.7 under Section 15(4) of the Patents Act 1977

(71) Applicant(s)
Simoco International Limited
(Incorporated in the United Kingdom)
PO Box 24, St. Andrews Road, CAMBRIDGE, CB4 1DP, United Kingdom

(72) Inventor(s)
Ian Vince McLoughlin
Robert James Chance

(51) INT CL[7]
G10L 19/06

(52) UK CL (Edition R)
H4R RPNR
U1S S2215

(56) Documents Cited
GB 2220330 A EP 0793218 A2 EP 0742548 A2
US 4653099 A

(58) Field of Search
UK CL (Edition R) H4R RPNR RPVS
INT CL[7] G10L 19/06
Online:WPI,EPODOC,JAPIO,INSPEC

(74) Agent and/or Address for Service
Frank B Dehn & Co
179 Queen Victoria Street, LONDON, EC4V 4EL, United Kingdom

(54) Abstract Title
Using LSP to alter frequency characteristics of speech

(57) A speech communication system comprises a receiving unit 14 which receives speech data and uses that data to output speech 15. The characteristics of the received speech data are altered by a processing unit 10 to make the speech more intelligible by altering line spectral pair data representing the speech to alter the frequency of a component in the speech spectrum. For example, an automatic examination of background noise amplitudes around the frequency of a formant of the speech might reveal that shifting the formant frequency upwards or downwards by 10% may improve intelligibility. If this is likely (perhaps because the noise amplitude reduces at a frequency 10% lower than the formant frequency), then the processing unit shifts the appropriate line spectral pair data by the corresponding amount.

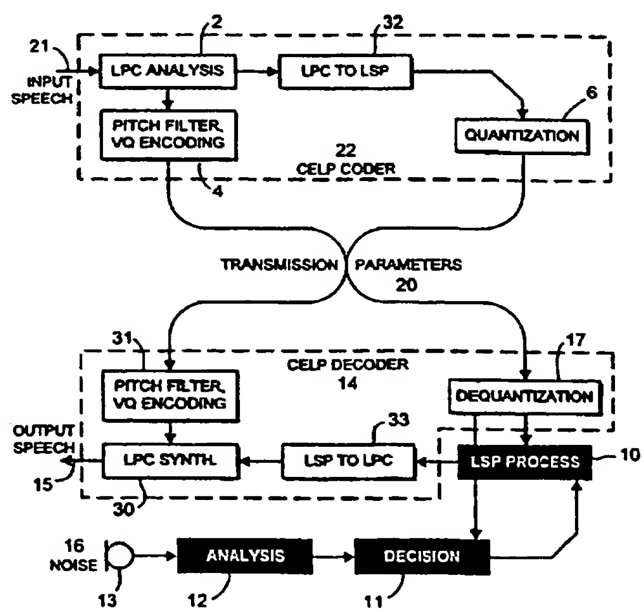

FIG. 2

GB 2 343 822 A

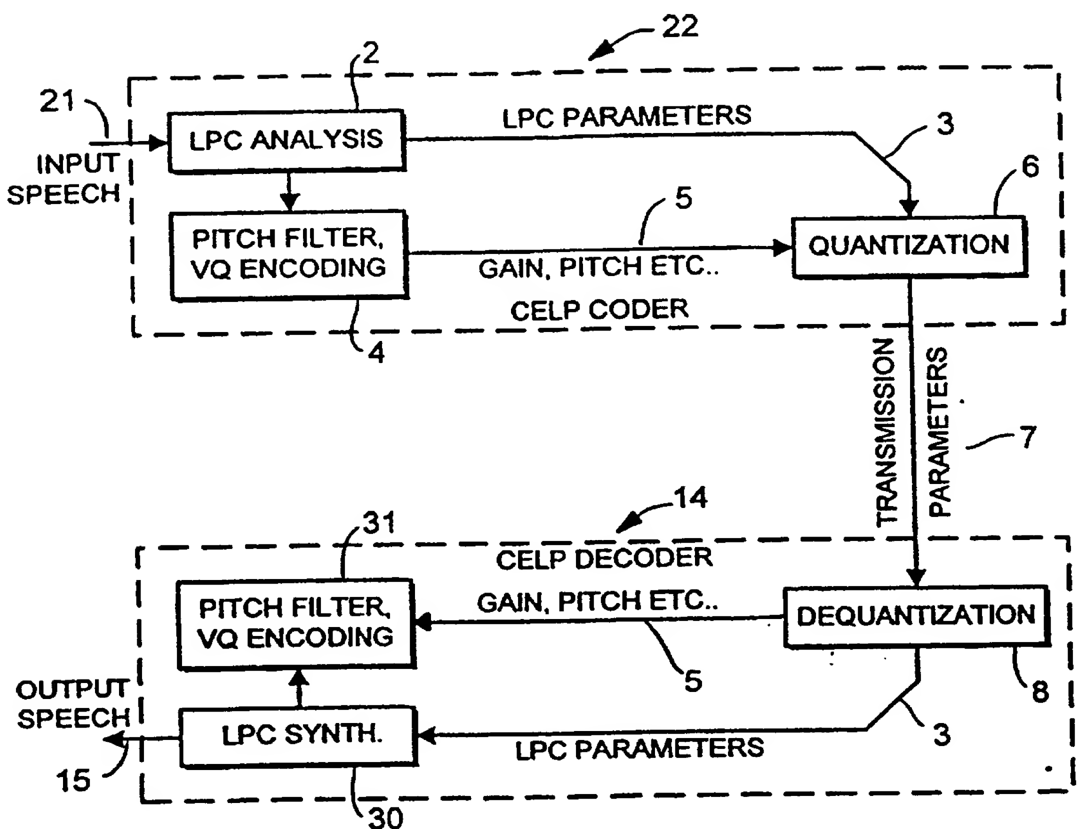
22
2
21
LPC ANALYSIS
LPC PARAMETERS
3
INPUT
SPEECH
6
5
PITCH FILTER,
VQ ENCODING
QUANTIZATION
GAIN, PITCH ETC..
CELP CODER
4
TRANSMISSION
PARAMETERS
7
14
31
CELP DECODER
PITCH FILTER,
VQ ENCODING
GAIN, PITCH ETC..
DEQUANTIZATION
5
8
OUTPUT
SPEECH
LPC SYNTH.
3
LPC PARAMETERS
15
30

FIG. 1

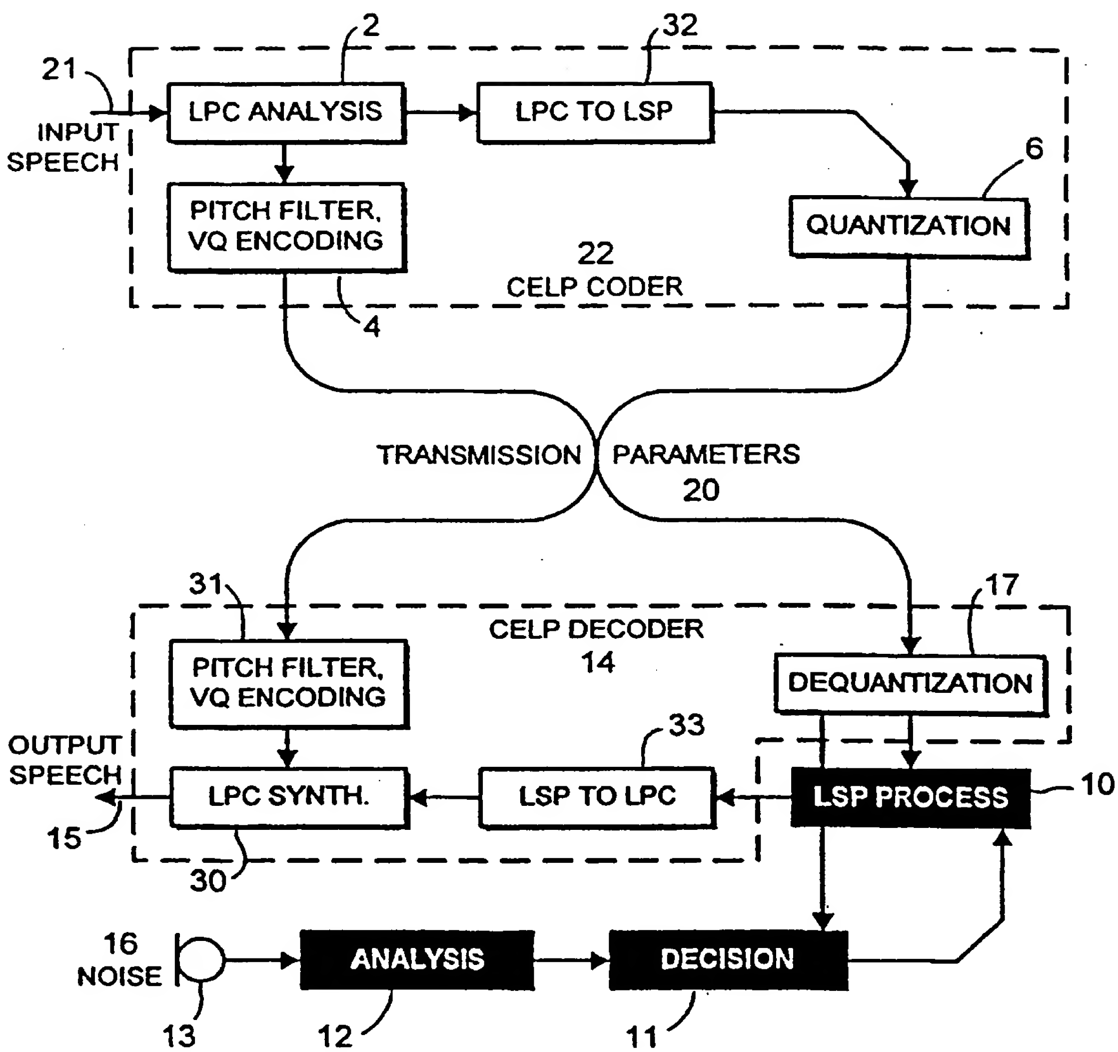
21
INPUT SPEECH
2
LPC ANALYSIS
32
LPC TO LSP
6
PITCH FILTER, VQ ENCODING
QUANTIZATION
22
CELP CODER
4
TRANSMISSION
PARAMETERS
20
31
17
CELP DECODER
14
PITCH FILTER, VQ ENCODING
DEQUANTIZATION
OUTPUT SPEECH
33
LPC SYNTH.
LSP TO LPC
LSP PROCESS
10
15
30
16
NOISE
ANALYSIS
DECISION
13
12
11

FIG. 2

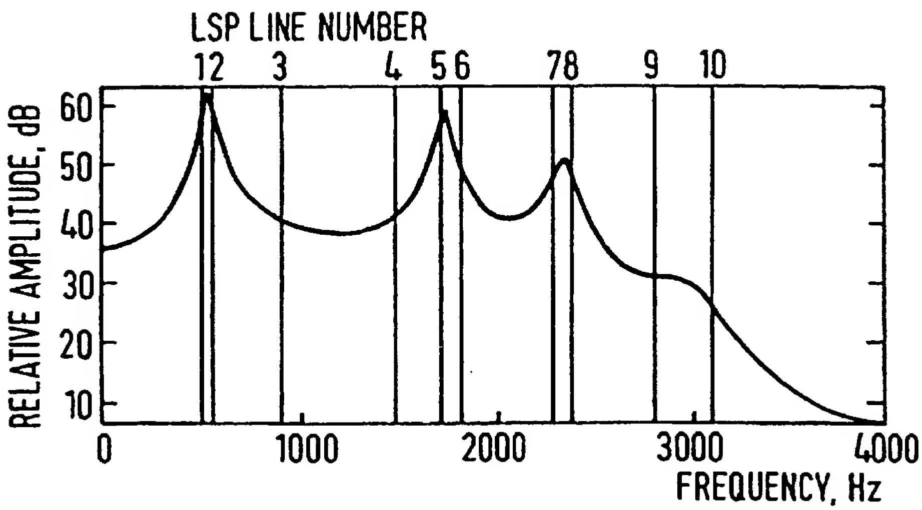
LSP LINE NUMBER
1 2 3 4 5 6 7 8 9 10
RELATIVE AMPLITUDE, dB
60
50
40
30
20
10
0
1000
2000
3000
4000
FREQUENCY, Hz

FIG. 3

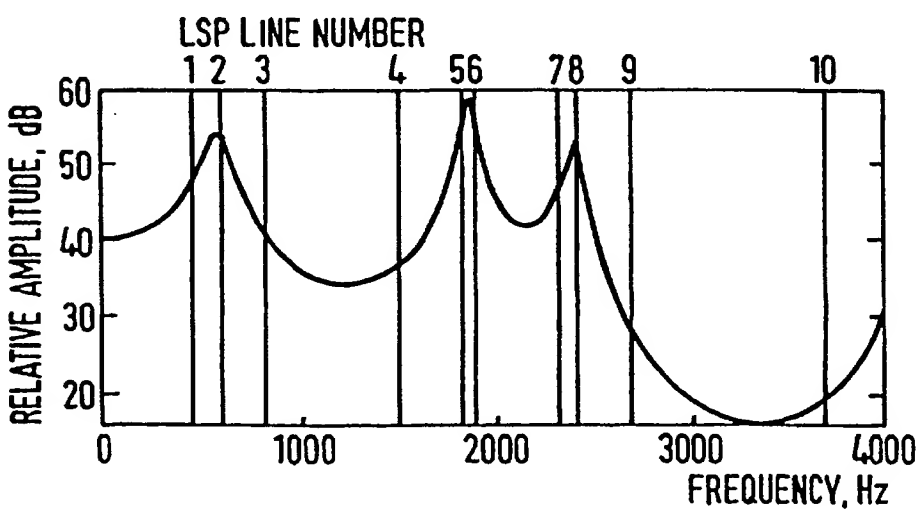
LSP LINE NUMBER
1 2 3 4 5 6 7 8 9 10
RELATIVE AMPLITUDE, dB
60
50
40
30
20
0
1000
2000
3000
4000
FREQUENCY, Hz

FIG. 4

METHOD AND APPARATUS FOR SPEECH ENHANCEMENT IN A SPEECH COMMUNICATION SYSTEM

The present invention relates to a method and apparatus for speech enhancement in a speech communication system, and in particular to such a method and apparatus for enhancing speech to make it more intelligible to a listener in a noisy environment.

Speech communication systems such as mobile phones and radios are often used in noisy environments, such as inside vehicles. Furthermore, this environmental noise can vary during a conversation. This varying environmental noise can make it very difficult for a listener to understand the speech being output by their phone or radio.

According to one aspect of the present invention, there is provided a method for increasing the intelligibility of speech output by a speech communication system to a listener using the system, comprising:

analysing the current background acoustic noise environment of the speech communication system;

determining using the results of the background noise analysis whether the speech to be output to the listener would be intelligible to the listener in the current background noise; and

altering the characteristics of the speech to be output by the speech communication system on the basis of said determination such that the altered speech output by the speech communication system has enhanced intelligibility to the listener in the current background noise.

According to a second aspect of the present invention, there is provided a speech communication system comprising:

means for analysing the current background acoustic noise environment of the speech communication system;

means for determining using the results of the background noise analysis whether speech to be output by the speech communication system would be intelligible to a listener in the current background noise environment; and

means for altering the characteristics of the speech to be output by the speech communication system to enhance the intelligibility of the speech to a listener in the current background noise in accordance with the output of said determining means.

The present invention thus monitors the background noise in which a speech communication system is being used (i.e. the external environmental acoustic noise in the vicinity of the listener) and can adjust the characteristics of the speech to be output by the speech communication system to the listener to make it more intelligible in that current background acoustic noise. It therefore provides enhanced intelligibility of speech output as sound by, for example, the loudspeaker or earpiece of a mobile phone or radio when used in noisy environments.

Furthermore, because the present invention analyses current background noise, it can take account of changes in the background noise and enhance the speech accordingly. In the present invention the background acoustic noise is therefore preferably continuously analysed and the speech continuously altered on the basis of that analysis. This provides for dynamic enhancement of the speech and is particularly advantageous in environments where background noise can change continuously and significantly, such as in a vehicle.

The background acoustic environmental noise can be analysed by various techniques, as is known in the art. It can be picked up or sampled using, for example, the

usual microphone for picking up the user's speech of the speech communication system (e.g. mobile phone or radio), or a separate microphone.

An example background noise analysis system would be a process whereby the user's speech (for example in the microphone signal) is detected (using one of many common techniques, such as adding all input noise values in a given time interval and comparing these against a threshold) and the acoustic background noise is analysed during the gaps between the speech periods.

The sampled noise would then be analysed (perhaps using linear prediction) to determine both its spectral content and its amplitude. LPC (linear prediction coefficient) values resulting from a linear predictive analysis contain sufficient spectral information, and a gain parameter could be used to relate the relative amplitudes of the LPC parameters to absolute amplitudes.

The intelligibility of speech to be output by the speech communication system in the current background noise can be determined using any known standard technique to determine whether the speech would be intelligible to an average listener in the current background noise (i.e. any suitable technique for assessing the effect of that noise on the listener's perception of the speech).

Preferably, descriptions of the speech and the background noise in the form of spectral analyses and amplitude scaling factor (gain) are compared to determine if the speech would be audible to a listener in that noise.

In a preferred embodiment the speech is first classified into two or more categories, and the amplitude of one of the speech categories at one or more frequencies compared with the noise amplitude at those frequencies.

In one such comparison process, the speech contents could initially be classified into non-speech, voiced

speech or unvoiced speech. If non-speech is present (perhaps a pause between words), then the audibility of this is unimportant and so it can be ignored.

If voiced speech is present, then its intelligibility needs to be determined. This is preferably done by comparing the amplitude of one or more, or most preferably each, spectral peak and/or of one or more, or most preferably each, formant (as is known in the art, voiced speech contains a series of resonant peaks at varying frequencies called formants which convey a great deal of information and to which spectral peaks in the spectral plot of the speech often correspond) in the voiced speech with the noise amplitude at the frequency of the peak or formant, respectively. If more than one peak or formant is to be considered, then the amplitude of each peak or formant should be compared with the noise amplitude at the frequency of the respective peak or formant.

Most preferably, the speech is determined to be unintelligible if the noise amplitude at any formant frequency or spectral peak or at a particular number of formant or spectral peak frequencies exceeds the corresponding formant or spectral peak amplitude(s).

Such comparison of the relative amplitudes of spectral peaks and formants in the speech with the background noise will give a good indication of the intelligibility of the speech, because it effectively determines the intelligibility of the speech in terms of a human listener model of intelligibility, i.e. it assesses the intelligibility of the speech in a manner that models closely a human listener's actual perception of the speech. As a well-known psycho-acoustic theory states, a sound of a given frequency will be masked by a second coincidental sound of similar frequency, and if the second sound is loud enough, then the former sound will be inaudible. Thus the Applicants have recognised that in the case of speech, loud noises with frequencies

similar to those of formants or spectral peaks in the speech will mask the speech. Thus comparison of the amplitude of one or more or each formant or one or more or each spectral peak in the speech with the noise amplitude at the corresponding frequency or frequencies will give a good indication of the audibility of that (or those) formant(s) or spectral peak(s) and thus of the intelligibility of the speech to a human listener.

Other speech classifications and categories could be used if desired. For example, the speech could be classified into vowel and consonant sounds (or other speech sounds). Preferably, a classification is used which is helpful or appropriate to determining intelligibility. Thus preferably, as in the above example, the classification includes a category which includes formants of the speech (preferably only formants) and that category is compared with the noise. Preferably the classification is into formant containing and non-formant containing categories.

Once the intelligibility of the speech has been determined, the speech can be altered to make it more intelligible in accordance with that determination. Preferably, if it is determined that the speech would be unintelligible, then the speech characteristics are altered, but not otherwise.

Alteration of the speech characteristics can be done in various ways, as is known in the art. It is preferably done by increasing the volume (amplitude) and/or altering the frequency of speech components and in particular the formants and/or spectral peaks in the speech.

In a particularly preferred such arrangement, the speech characteristics will be altered by adjusting the positions of the formants and/or spectral peaks in the speech spectral plot. Such alterations will have a more perceptible effect on the speech to a human listener and thus are particularly effective for increasing the

intelligibility of the speech. For example, one or more peaks or formants could be shifted upwards or downwards in frequency, or the amplitude of one or more peaks or formants could be increased (corresponding to a decrease in bandwidth), or the bandwidth of one or more of the peaks or formants could be increased (corresponding to a decrease in amplitude).

Thus, for example, the volume of the formants can be increased such that they are audible over the background noise. However, this can be an undesirable way of altering the speech characteristics as speech volume levels sufficient to cause hearing loss (if sustained) may be required to make the speech intelligible in certain situations, notably those within noisy motor vehicles.

Preferably therefore the frequency of speech components such as formants or peaks in the speech spectrum is adjusted. This is preferably done to move them to a frequency where the noise level is lower, such that the components, e.g. peaks or formants, are audible (i.e. have an amplitude greater than the noise) at that frequency.

The alteration of speech characteristics is preferably carried out in accordance with the results of the analysis of the background noise, and may be dependent upon the present or past values of the noise. Using present values of noise, a direct comparison may be made and an alteration made to the speech characteristics; using past values, it is possible to make predictive changes. For example, if the noise analysis indicates the noise amplitude reduces at a particular frequency to a level at which a presently inaudible formant would be audible, the speech characteristics could be altered to change the frequency of that formant to that particular frequency.

The actual alteration of speech characteristics can be carried out in a number of ways, as is known in the

art. For example, the speech signal could be passed through an adaptive filter, such as a perceptual error weighting filter (as described in CHEN, J. H., COK, E.V., LIN, Y., JAYANT, N., and MIECHER, M.J., "A low delay CELP coder for the CCITT 16 kb/s speech coding standard". *IEEE J. Scl. Ateas Commun*. 1992, 10. (5). pp 830-849) to narrow or widen the formant bandwidth. Alternatively the amplitude peaks could be clipped so that the energy in the unvoiced parts of the speech becomes a more significant part of the total speech energy. This can increase intelligibility but at the expense of sound quality.

In a particularly preferred embodiment, the speech characteristics are altered by altering line spectral pair (LSP) data representing the speech.

As is known in the art, line spectral pairs are representations of the linear-prediction parameters derived for periods of sound. Where the sound is speech, the resonant frequencies in the speech or formants, can be noted in the linear-prediction spectrum. LSP values usually uniquely relate to positions of such resonances or formants in the linear-prediction spectrum. Thus LSP data can be used to represent speech, and the Applicants have recognised that by altering the LSP data, characteristics such as the frequency and amplitude of formants in the speech can be adjusted. This allows the speech characteristics to be adjusted relatively easily and in a way that can readily change the speech as perceived by a listener and at a much lower computational overhead than when using, for example, adaptive filtering. Also, such adjustment does not eliminate parts of the speech spectrum, but rather modifies them.

Furthermore, many speech communication systems such as speech coding/decoding systems used in mobile telephones or modern digital radio systems, utilise a linear-prediction model of speech, and convert this to

an LSP representation for transmission. The LSP representation is generally used within such speech systems for reasons of information security and transmission efficiency.

Thus this embodiment of the present invention is particularly advantageous in such systems which use LSPs for speech transmission, since the LSP information that is transmitted may be altered in the speech communication system when it is received to enhance the intelligibility of the speech. This altered LSP data would then be converted back to linear-prediction parameters and hence reconstructed into speech and output as sound, but with altered characteristics.

It is believed that the adjustment of LSPs representing speech in a speech communication system to change the characteristics of speech output by that system could be advantageous in itself.

Thus according to another aspect of the present invention, there is provided a method of altering the characteristics of speech to be output to a listener in a speech communication system in which the speech data to be processed and output by the speech communication system includes line spectral pair data, comprising altering the line spectral pair data in the speech data.

According to a further aspect of the present invention, there is provided a speech communication system in which the speech data to be processed by the speech communication system includes line spectral pair data, comprising means for altering the line spectral pair data in the speech data processed by the speech communication system to change the characteristics of the processed speech as heard by a listener.

In these aspects of the invention, the alteration of the LSP data in the speech data is preferably used for the purpose of enhancing the intelligibility of the output speech when listened to in a noisy environment (but it could be useful in other situations where it is

desired to alter the characteristics of speech as heard by a listener, e.g. to disguise the speaker's voice). Thus these aspects of the present invention preferably comprise the technique of adjusting the values of LSPs found within the speech data based upon an analysis of the background acoustic noise environment of the system (i.e. the listener). Preferably, the frequency or the power and bandwidth of specific frequency-domain features, such as formants, found in the speech are altered in this way.

The LSP alterations can be designed to affect the reconstructed speech in specific ways and in particular to enhance the intelligibility of the speech over the background noise, as discussed above. For example, the particular line spectral pair (LSP) associated with a formant can be identified and its separation (or spacing) then widened or narrowed to increase or decrease the formant bandwidth. Alternatively or additionally, line spectral pairs can be moved higher or lower in frequency to increase or decrease the frequency of particular formants.

The LSP information is preferably altered by adding or subtracting values to one or more LSPs (or LSP lines), or by moving one or more LSPs (or LSP lines) in the speech spectrum. The values may be determined in accordance with the analysis of the background noise, and may be dependent upon the present or past values of each LSP. Using present values of LSP data, a direct comparison can be made with the ambient noise and an adjustment made to the LSP data; using past values, it is possible to make predictive changes.

In a particularly preferred such arrangement, the invention includes making a numerical increment or decrement in the value of any or all of the set of LSPs (or LSP lines) defining the speech. Thus individual or groups of LSPs can be moved to: shift one or more spectral peaks or formants in frequency (either upwards

or downwards); or change the amplitude (either to increase the amplitude (decrease the bandwidth) or decrease the amplitude (increase the bandwidth)) of one or more spectral peaks or formants.

For example, the separation between the values of two or more of a set of LSP lines (and most preferably between a pair of LSP lines) can be narrowed or widened to narrow or widen frequency features (such as spectral peaks or formants) found in the speech frequency spectrum. Alternatively or additionally, the values of two or more of a set of LSP lines (and most preferably of a pair of LSP lines) can be incremented or decremented, most preferably by identical amounts (either in absolute terms or as a percentage of their original values), to adjust the centre frequency of features (such as spectral peaks or formants) found in the frequency spectrum of the speech.

In a particularly preferred embodiment, line spectral pairs are translated in frequency so as to change the centre frequency of particular peaks or formants in the speech data. As discussed above, this is a particularly advantageous way of changing speech characteristics as heard by a listener, for example to increase intelligibility over background noise.

It is also possible to predict the behaviour of the background noise from an analysis of previous changes in its spectral content, to enable a faster or more appropriate adjustment to the LSPs. This is particularly applicable to repetitive noise such as a siren in a police car, fire appliance or ambulance. Knowledge of which way the frequency of the interfering noise is changing may affect the decision about which way to shift the formant frequencies.

Any or all of the above adjustments can be used individually or in combination to alter the speech characteristics of the speech to be output by the speech communication system in accordance with the analysis of

the background noise of the listener to make the speech output by the speech communication system more intelligible to the listener.

The present invention has been described in relation to speech communication systems, such as mobile phones and radios. It is particularly suited to use in speech decoders, such as would be found for example in mobile phones or mobile radios. However, it would also be applicable (and in particular the aspects relating to LSP alteration would be applicable) to use in speech coders where it was desired to alter the characteristics of the user's input speech to be transmitted by the speech coder (for example to increase intelligibility over the speaker's background noise). It would also be applicable in radio receivers, televisions, or other devices which broadcast speech to listeners. Also although it has been described with particular reference to increasing the intelligibility of speech, it could also be used to increase the intelligibility of other sounds, such as music.

A preferred embodiment of the present invention will now be described by way of example only, and with reference to the accompanying drawings, in which:

Figure 1 shows a generic CELP codec structure;

Figure 2 shows a block diagram of a typical speech communication system in accordance with the present invention;

Figure 3 shows the frequency spectrum of a period of sound, with numbered LSP values for that sound overlaid as vertical lines; and

Figure 4 shows the frequency spectrum of a period of sound derived from the LSP values of Figure 3 with specific alterations. The altered LSP values for that sound are overlaid as vertical lines.

The present invention is particularly applicable to use in a speech codec system such as would be used in a mobile phone or radio system. An example of such a

codec structure is shown in Figure 1, in the form of a generic CELP coder.

The general CELP (codebook-excited linear prediction) structure was introduced in 1985 (see, for example, Shroeder MR, Atal BS, "Code-excited linear prediction (CELP): high-quality speech at very low bit rates", ICASSP, pp. 937-940, 1985), and many modifications have been made since.

A generic CELP codec structure 22 is shown in Figure 1. Figure 1 shows input speech 21 being analysed by linear prediction analyser unit or device 2 resulting in linear prediction (LPC) parameters 3. The remainder of the input signal which linear prediction cannot describe is passed to a pitch filter, VQ encoding block 4 which produces parameters representative of, for example, the gain and pitch of the speech. These processes are unimportant to the invention and vary widely between different CELP implementations in their detail, however they result in various other parameters which, together with the LPC parameters, describe the input speech.

The LPC parameters 3 and any other parameters (such as gain and pitch) 5 describing the input speech are quantized by a quantizer 6 and transmitted (as transmission parameters 7) to the CELP decoder 14 which dequantizes them using a dequantizer 8. These dequantized values are then used to recreate speech 15 to be output as sound to a listener. (The dequantizer 8 reproduces the LPC parameters 3 and other parameters 5 by means of an LPC synthesiser 30 and pitch filter, VQ decoding block 31, respectively, which reproduce the speech for it to be output as sound 15.)

LPC parameters may alternatively be converted to a different form prior to quantization in the coder (and also converted back to LPC coefficients after dequantization). Such forms may include log area ratios, PARCOR (reflection coefficients) and line

spectral pairs.

Differences in the representation of LPC parameter used and the types of (or usage of) pitch filter and vector quantizer (VQ) have led to many CELP variants. A small selection of examples are: MELP (mixed excitation linear prediction); VSELP (variable slope excitation linear prediction); SB-CELP (sub-band CELP); LD-CELP (low delay CELP); RELP (residual excitation linear prediction); RPE-LP (residual pulse excitation linear prediction); and others.

As noted above, in many such codecs the LPC parameters are transmitted as LSPs.

The terminology 'LSPs' refers to the parameters generated by a conversion of linear prediction coefficients using the line spectrum pair approach as described in the paper by Sugamura and Itakura (Sugamura N, Itakura F, "Speech analysis and synthesis methods developed at ECL in NTT - from LPC to LSP - ", Speech Communication, vol. 5, pp. 199-213, 1986). The linear prediction coefficients themselves are generated by any of the well-established analysis methods operating on a set of data (speech) such as those described in Makhoul J, "Linear prediction: a tutorial review", Proc. IEEE, vol 63, no. 4, pp. 561-580, 1975.

LSPs are generated via a mathematical transformation from LPCs and thus have identical information content, but different form. Many other mathematical transformations from LPCs have been determined, but none of the resulting parameters can be altered in the same way as LSPs and as described in the present invention.

The line spectral pair parameters may be referred to as line spectral frequencies, however this term is not applied exclusively to LSPs.

Mathematically speaking, LSP parameters may be defined as: the roots of the two polynomials formed by a particular re-arrangement of the coefficients of the

inverse linear prediction polynomial. These two polynomials may be called P and Q and are formed using the set of linear prediction coefficients, Ap (where p is the index of the array, usually running from 0 to the filter order, p), having the following recursive relationship:

$$P(z^{-1}) = A_p(z^{-1}) - z^{-(p+1)}A_p(z)$$
$$Q(z^{-1}) = A_p(z^{-1}) + z^{-(p+1)}A_p(z)$$

The roots obtained by solving the polynomials P and Q give the line spectral frequency parameters, referred to as line spectral pairs. Many methods exist to determine these roots, as explained in, for example, the paper by Sugamura and Itakura referred to above. The choice of method is irrelevant for the purposes of the present invention.

The set of LSPs are often scaled. With reference to a 'basic' LSP value, the cosine or sine of these are also referred to as LSPs. In addition, the basic LSP may reside in one of various domains, i.e. its maximum and minimum values may be between 0 and π, between 0 and 4000Hz (a typical sampling frequency), or within other arbitrary ranges such as 0 to 1.

As an aid to understanding of the present invention, a non-mathematical description of line spectral pairs (LSPs) will also be considered. As LSPs are derived from LPC and reflection coefficients, it is necessary to cover these first.

Linear prediction is the usage of a fixed-length formula to model an unknown system. The formula structure is fixed but the values to be inserted into the formula must be found. Linear predictive analysis is the process of finding the best set of values for that formula. These values are the linear prediction coefficients, and the best set of these values is the set that causes the equation output to resemble the

output of the system to be modelled most closely, when the inputs to the two systems are identical.

If the equation of that formula is re-ordered mathematically then another standard equation can be arrived at. The coefficients for the new equation are called reflection coefficients and can be found easily from the LPC coefficients.

The reflection coefficient equation is very easy to relate to a real system. For speech processing, the LPC analysis is attempting to find the best parameters that model a short period of speech. In physical terms, the model is made up of a number of different width but equal length tubes connected in series. The reflection coefficients fit well into this physical model as the reflection coefficients relate directly to the difference between each consecutive tube.

When air is blown down tubes, resonances occur (organ pipes). In a human vocal tract, air originates at the glottis (which opens and closes rapidly) and proceeds through the vocal tract to be expelled at the mouth. The sound relates strongly to the shape of the vocal tract due to the resonances.

The LSP parameters each relate to the resonant frequency of one of the connected tubes. Half of the parameters are generated assuming that the source end of tubes is open, and half assuming that it is closed. In fact, the glottis opens and closes rapidly and so is neither open nor closed. Thus each true spectral resonance occurs between two nearby line spectral frequencies and these two values are considered to be a pair (thus line spectral pair).

An embodiment of the present invention in a speech communication system comprising a speech codec, and using LSP alteration to enhance the intelligibility of speech in a noisy environment is shown in Figure 2, and the signal processing is illustrated in Figures 3 and 4. The system as shown in Figure 2 has many features in

common with the system of Figure 1 and thus the same reference numerals have been used for the like features of the systems.

The LSP alteration mechanism may act within a speech codec (a codec comprises both a coding 22 and a decoding 14 mechanism) in the positions shown in Figure 2 (i.e. in the speech decoder 14). The speech coder 22 transforms the input speech 21 into a set of condensed parameters 20 suitable for transmission by radio or other means to a receiving unit 14. (It should be noted that in this arrangement the LPC parameters produced by the linear prediction analyser 2 are converted to line spectral pair data by an LPC to LSP converter 32 before being quantized by the quantizer 6.) The receiving unit then decodes the transmitted data to reconstruct speech 15. By way of example, the coding unit 22 may reside in an office telephone and the decoding unit 14 within a mobile telephone handset.

In this embodiment alterations to the data received by the decoding unit, where that data comprises LSP information, are performed. This alteration unit is shown in Figure 2 as LSP processor 10.

The LSP processing depends upon the degree and type of acoustic noise background 16 that is present in the environment of the listener. The analysis unit 12 shown in Figure 2 determines the type and level of background noise by use of a microphone 13 which picks up, *inter alia*, the actual external background acoustic noise of the listener's environment.

An example of a noise analysis system would be a process whereby the user's speech is detected (using one of many common techniques, such as adding all input noise values in a given time interval and comparing these against a threshold) and the external acoustic background noise is considered during the gaps between speech periods.

The sampled noise must then be analysed (perhaps

using linear prediction) to determine both its spectral content and its amplitude. LPC (linear prediction coefficient) values resulting from a linear predictive analysis contain sufficient spectral information, and a gain parameter would relate the relative amplitudes of the LPC parameters to absolute amplitudes.

The decision device or unit 11 determines whether the speech data currently being received by the decoder and replayed as sound via the loudspeaker or ear piece of the mobile telephone unit would be intelligible to an average listener in the current background acoustic noise 16 of the mobile telephone unit (i.e. listener).

If the decision unit determines that speech is readily intelligible then no processing is necessary and the processing unit 10 would not alter the dequantized LSP parameters 17 which have been passed to it by the standard speech decoder, before passing them to the LSP to LPC converter 33.

On the other hand, if the decision unit determines that the speech is unintelligible, then processing is necessary and the processing unit 10 would alter the dequantized LSP parameters to alter the speech characteristics before passing them to the LSP to LPC converter for subsequent playback to the listener. The decision unit may also predict that the speech will shortly become unintelligible.

Inputs to the decision process are descriptions of speech and background noise, in the form of spectral analyses and amplitude scaling factor (gain). It is necessary to compare the speech and noise data to determine if the speech would be audible to a listener in that noise.

Comparison could be to initially classify the contents of the speech signal into non-speech, voiced speech or unvoiced speech. If non-speech was present (perhaps a pause between words), then the audibility of this is unimportant and thus no enhancement is required,

and the LSP-process module would be commanded to perform no processing.

If voiced speech is present (voiced speech contains a series of resonance peaks at various frequencies called formants), then the amplitude of each formant would be compared to the noise amplitude at that frequency to determine its audibility. If the noise amplitude at any formant frequency exceeds the formant amplitude then formant adjustment is required.

Other known techniques for determining the intelligibility of the speech to be output could be used, if desired.

The LSP process unit 10 performs mathematical operations on individual LSPs to enhance the speech under the control of the decision unit.

The exact operations would depend upon the directions of the decision process. One speech enhancement function would entail the shifting of LSP lines to more favourable locations.

For example, an automatic examination of the noise amplitudes around the formant frequency might reveal if, perhaps, shifting the formant frequency upwards or downwards by 10% may improve matters. If this is likely (perhaps because the noise amplitude reduces at a frequency 10% lower than the formant frequency), then the LSP processing block is directed to shift the appropriate LSPs by the corresponding amount.

If, for example, the formant that requires moving is located at 600Hz, then two LSP coefficients would exist, usually very close to and either side of 600Hz. If audibility is to be improved by a downwards shift of 10%, then the values of these two LSP parameters would each be multiplied by 0.9 to effect that shift. The LSP adjustment itself is confined to within the LSP process block.

As a further example, if the decision module determined that shifting lines 1 and 2 from a set of

LSPs downwards in frequency by 10% would improve intelligibility, then the values of lines 1 and 2 would both be multiplied by a factor of 0.9.

If the decision module determined that upward shifting of line 3 by 100Hz improves intelligibility then an amount would be added to line 3. This amount would be equal to 100 if the LSP parameters were scaled to have values in Hz, or would more generally be

$$\frac{100 \times 2\Pi}{f_s}$$

where f_s is the sampling rate of the system, and the values of the LSPs are confined to the angular frequency domain.

Other types of processing are possible, but may all be described as adding/subtracting values to one or more LSP lines (with adding LSP lines to themselves being equivalent to multiplication). The values may be determined by the decision module or may be dependent upon the present or past value of each LSP line.

An example of such LSP processing is illustrated in Figure 3, in which the frequency spectrum of a period of sound has been plotted, and the 10 LSP lines obtained from analysing this sound have been overlaid. LSP values may be readily converted to and from the LPC parameters from which the spectrum is plotted. For the specific example in question, Figure 3 thus shows the frequency spectrum of the sound obtained from the analysis of speech 21 in the CELP coder 22 of Figure 2.

In the case of a standard CELP decoder, operating without the benefit of this invention, the output speech 15 would be reconstructed using the data of Figure 3. When the invention is included, the LSP processing block 10 would be capable of altering the LSP values in order to change the output speech 15.

For the specific example of Figure 4, certain of the LSP values of the spectrum of Figure 3 have been altered and a new set of LPC coefficients have thus been generated forming the spectrum as shown in Figure 4. Referring to the LSP values of the original spectrum in Figure 3, three operations have been performed:

1. The separation between lines 1 and 2 has been increased by moving both of the lines further apart (in other words 1 has been lowered in frequency and 2 has been raised)

2. Lines 5 and 6 have been increased in frequency

3. Line 10 has been increased in frequency.

The three actions have specific consequences to the sound that is transmitted:

1. Lines 1 and 2 lie on either side of a spectral peak. The movement in the two lines has induced this spectral peak to both reduce in amplitude and become wider (equivalent to an increase in bandwidth).

2. Lines 5 and 6 lie on either side of a second spectral peak. The movement of these two lines has induced that peak to increase in frequency.

3. Line 10 previously lay to the right of a very small spectral 'bump' which is now no longer evident as the line has been increased in frequency by a substantial amount.

In this specific example of a speech codec, the sound under analysis is speech. The spectral peaks evident in the spectral plots will then often, as

discussed above, correspond to formants, important constituents of speech that convey a great deal of information. The LSP-based adjustments discussed above have thus changed the characteristics of the speech to be output to and as it will be perceived by the listener. For example, in the case of vowels, moderately widening the lines corresponding to spectral peaks (i.e. increasing the bandwidths of the formants) has been found to improve intelligibility.

The example shown in Figure 2 additionally analyses the noise present in the environment of the listener to determine if the speech to be replayed to that listener is intelligible. If not, then speech characteristics are altered in the present invention to improve the intelligibility of the speech by the operation of moving individual or groups of LSPs to provide the following set of operations:

1. Shift peak/formant upwards in frequency.

2. Shift peak/formant downwards in frequency.

3. Increase amplitude (decrease bandwidth) of peak/formant.

4. Increase bandwidth (decrease amplitude) of peak/formant.

A well-known psychoacoustic theory states that a sound of given frequency will be masked by a second coincidental sound of similar frequency. If the second sound is loud enough, then the former sound will be inaudible. Thus, in the case of speech, the Applicants have recognised that loud noises with frequencies similar to those of the formants will mask the speech. In order to hear the speech it is necessary to either increase the volume or alter the frequency of the speech

components.

Volume alteration is relatively straightforward, but it should be noted that speech volume levels sufficient to cause hearing loss (if sustained) may be required to make speech intelligible in certain situations, notably those within noisy motor vehicles. It is therefore preferred to alter the frequency of speech components.

As can be seen, the present invention offers a method of reducing the masking of speech by acoustic background noise (and thus improving intelligibility) through an efficient process that may be combined with many of the current standard mobile telephone and radio systems, and standard speech codecs in such systems.

Speech enhancement results when an analysis of the listener's background noise environment is combined with corrective LSP alteration, which adjusts received transmitted speech data to be replayed to the listener in order to improve the chances of the listener hearing the processed sounds. The technique adjusts the values of LSPs found within the speech data codec based upon an analysis of the background acoustic noise environment of the listener. Preferably, the frequency or the power and bandwidth of specific frequency-domain features found in the received speech are altered in this way.

Claims

1. A method of altering the characteristics of speech to be output to a listener in a speech communication system in which the speech data to be processed by the speech communication system and output as sound includes line spectral pair data, comprising altering the line spectral pair data in the speech data to alter the frequency of a component in the speech spectrum.

2. A method as claimed in claim 1, wherein the frequency of a formant in the speech spectrum is altered.

3. A method as claimed in claim 1 or claim 2, wherein the centre frequency of a spectral peak in the speech spectrum is altered.

4. A method as claimed in any one of claims 1 to 3, wherein the line spectral pair data is altered by changing the frequency of a line spectral pair in the speech spectrum.

5. A method as claimed in any one of claims 1 to 4, wherein the line spectral pair data is altered by incrementing or decrementing a pair of line spectral pair data lines by identical amounts.

6. A method as claimed in any one of claims 1 to 5, further comprising altering the line spectral pair data by decreasing the spacing of a line spectral pair in the speech spectrum.

7. A speech communication system in which the speech data to be processed by the speech communication system includes line spectral pair data, comprising means for altering the line spectral pair data in the speech data

processed by the speech communication system in such a manner that the frequency of a component in the speech spectrum is changed to change the characteristics of the processed speech as heard by a listener.

8. A system as claimed in claim 7, wherein the means for altering the line spectral pair data comprises means for altering the frequency of a formant in the speech spectrum.

9. A system as claimed in claim 7 or claim 8, wherein the means for altering the line spectral pair data comprises means for altering the frequency of a spectral peak in the speech spectrum.

10. A system as claimed in any one of claims 7 to 9, wherein the means for altering the line spectral pair data comprises means for altering the frequency of a line spectral pair in the speech spectrum.

11. A system as claimed in any one of claims 7 to 10, wherein the means for altering the line spectral pair data comprises means for incrementing or decrementing a pair of line spectral pair data lines by identical amounts.

12. A system as claimed in any one of claims 7 to 11, wherein the means for altering the line spectral pair data further comprises means for decreasing the spacing of a line spectral pair in the speech spectrum.

The Patent Office

INVESTOR IN PEOPLE

25

Application No:	GB 0001586.7	**Examiner:**	B.J.SPEAR
Claims searched:	1-12	**Date of search:**	13 March 2000

Patents Act 1977
Search Report under Section 17

Databases searched:

UK Patent Office collections, including GB, EP, WO & US patent specifications, in:

UK Cl (Ed.R): H4R (RPNR,RPVS)

Int Cl (Ed.7): G10L 19/06

Other: Online:WPI, EPODOC, JAPIO, INSPEC

Documents considered to be relevant:

Category	Identity of document and relevant passage		Relevant to claims
X	GB2220330A	(British Telecommunications) Whole document, eg Abstract, Fig. 7B , page 3 lines 18-20, page 17 line 6 to page 18 line 10, and claim 12	1 and 7
X	EP0793 218A2	(Sony) Whole document, eg Abstract, Fig. 3, and page 3 ll 24-29.	1 and 7
X	EP0742548A2	(Mitsubishi)Whole document, eg Abstract, Fig. 1, page 8 line 10 to page 10 line 37 and claim 1.	1-4,7-10
X	US4653099	(Casio)Whole document, eg Abstract and claim 1. Against claims 1 and 7.	1 and 7

X	Document indicating lack of novelty or inventive step	A	Document indicating technological background and/or state of the art.
Y	Document indicating lack of inventive step if combined with one or more other documents of same category.	P	Document published on or after the declared priority date but before the filing date of this invention.
		E	Patent document published on or after, but with priority date earlier than, the filing date of this application.
&	Member of the same patent family		